**Final Report: Smart City Public Transportation Project**

Optimizing Smart City Public Transportation Through Data and Technology

DS-670-HYB2-23WNTR

Capstone: Big Data & Bus Analysis

Reda Mastouri

30/01/2024

Kethana devi kadiyala

Venkata Sairam Rohith Mechineni

Deva Sai Vikas Vakkalagadda

*Executive Summary:*

Our Smart City Public Transportation project is a groundbreaking initiative aimed at revolutionizing urban transportation systems. This comprehensive report delves into the core problem statement, provides a detailed overview of our innovative solution, and offers in-depth explanations of the code implementation, drawing extensively from the solution report.

*Problem Statement:*

In the urban transportation landscape, challenges such as congestion, environmental sustainability concerns, and a lack of real-time information have emerged as critical obstacles. Our project addresses these issues by specifically focusing on the lack of real-time information in existing public transportation systems, aiming to reduce inefficiencies, minimize commute times, and enhance overall commuter satisfaction.

*Solution Overview:*

Our multifaceted solution comprises exploratory data analysis, various clustering techniques (KMeans, Hierarchical, DBSCAN, GMM, Agglomerative), and the development of a Flask application. The project's high-level architecture integrates data collection, processing, communication, decision support, user interface, and security layers.

*Code Implementation Explanations:*

1. **Exploratory Data Analysis and Clustering:**
   * *Data Preprocessing:* Explored the dataset, removed columns with missing values, and imputed values.
   * *KMeans Clustering:* Determined optimal clusters using the Elbow method, evaluated quality with silhouette score.
   * *Hierarchical Clustering:* Visualized hierarchical clustering through a dendrogram, determined optimal clusters.
   * *DBSCAN:* Implemented DBSCAN for spatial clustering, evaluated the silhouette score for effectiveness.
   * *Gaussian Mixture Model (GMM):* Applied GMM for data distribution modeling, assessed the silhouette score for quality.
   * *Agglomerative Clustering:* Employed Agglomerative Clustering, visualized spatial distribution of clusters.
2. **Flask Application (app.py):**
   * *Data Integration:* Loaded preprocessed dataset into the Flask app for real-time predictions.
   * *GMM Integration:* Integrated the trained GMM model into the Flask application for dynamic clustering.
   * *Real-time Plotting:* Utilized seaborn and matplotlib for real-time plotting of clusters on geographical maps.
3. **User Interface (index.html):**
   * *Form Design:* Developed a clean and intuitive HTML form with input fields for user-friendly interaction.
   * *Cluster Prediction:* Leveraged Flask to process user inputs, predict the cluster using the GMM model, and displayed results visually on the geographical map.
4. **Testing and Iteration:**
   * *Rigorous Testing:* Conducted extensive testing to ensure accuracy and reliability of cluster predictions.
   * *Performance Optimization:* Explored opportunities to optimize the application's performance for responsiveness with larger datasets.
   * *Iterative Refinement:* The application underwent iterative refinement based on user feedback and testing outcomes, ensuring continuous improvement.

*Future Steps:*

Outlined in the solution report, future steps include model refinement, feature enhancement, and continuous user feedback. The project aims to refine clustering models based on additional data, explore advanced features such as dynamic route planning, and prioritize user experience in further iterations.

*Conclusion:*

The Smart City Public Transportation project stands as a testament to our commitment to creating efficient, sustainable, and user-friendly urban transportation systems. By addressing key challenges and leveraging advanced technologies, our project contributes to the realization of smarter cities and an enhanced quality of life for residents.

**Instructions for Running the Flask App:**

1. Ensure Python is installed on your system.
2. Install required libraries using **pip install flask pandas matplotlib seaborn scikit-learn**.
3. Download the Flask app code (**app.py**, **index.html**) and the traffic dataset.
4. Navigate to the project directory in the terminal.
5. Run the Flask app with the command **python app.py**.
6. Access the app in your web browser at **http://127.0.0.1:5000/**.
7. Enter values in the input form, click "Predict Cluster," and explore the results.

**Final Result:**

